DOCUMENT RESUME

ED 322 430 CG 022 702

AUTHOR Hoover, Donald R.
TITLE Predicting the College Attendance Rate of Graduating High School Classes.
PUB DATE 90
NOTE 10p.
PUB TYPE Journal Articles (080) -- Statistical Data (110) -- Reports - General (140)
JOURNAL CIT CACD Journal; v10 1989-90

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Attendance; *College Bound Students; College Freshmen; Educational Attainment; *Enrollment Rate; Evaluation Methods; Higher Education; High Schools; *High School Seniors; *Predictive Measurement; *Predictor Variables; Student Educational Objectives

ABSTRACT

An important element of school counseling is providing assessments on the collective future needs and activities of a graduating school class. The College Attendance Rate (CAR) is defined here as the proportion of seniors graduating from a given high school, during a given year, that will enroll full-time at an academic college sometime during the following year. Estimated CARs of graduating high school classes are used by college and military recruiters and some education departments to plan and assess recruit market size. Most states keep annual records of CAR estimates. Three currently used methods of estimating CAR use: student intentions; educated guesses provided by high school officials; and linear regression based on high school and student body characteristics. Reasonably accurate CAR estimates can be produced by using educated guesses by high schools and by using student intentions. The third method, linear regression with high school demographic characteristics, has revealed interesting relationships between the variables and CAR but has demonstrated poor predictive ability. The predictive ability of linear regression can be greatly improved by using Scholastic Aptitude Test or American College Test scores and percentages of the graduating class taking college entrance examinations as predictor variables. (ABL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Predicting the College Attendance Rate of Graduating High School Classes

Donald R. Hoover

ED322430

CG 022702

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard Hoover
Donald R. Hoover

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Predicting the College Attendance Rate of Graduating High School Classes*

Donald R. Hoover

Abstract

Estimated college attendance rates (CARs) of graduating high school classes are used by college and military recruiters and some education departments to plan and assess. Most states keep annual records of CAR estimates. Three different methods of CAR estimation are compared and simple mathematical adjustments to improve CAR estimates are given.

An important element of school counseling is providing assessments on the collective future needs and activities of a graduating school class. The College Attendance Rate (CAR) is defined here as the proportion of seniors graduating from a given high school, during a given year, that will enroll full-time at an academic college sometime during the following year. CAR estimates have many important uses. For example, school districts and high school officials utilize CAR estimates as a tool in setting educational and counseling goals and allocating resources. College and military recruiters use CAR estimates to calculate high school recruit market size and evaluate recruiting efforts (Hoover, 1988). Relocating families and companies often use CAR estimates to measure desirability of high schools and school districts.

In the 30 or so states and territories (Hoover, 1988) that maintain annual records of estimated high school CAR, counselors are usually responsible for developing the estimates. Many states use educated guesses of CARs made by the counselors. Others base CAR estimates on student intentions (responses graduating seniors make to college intention questionnaires). While California currently makes no statewide effort to obtain high school CAR estimates, military and college recruiters often request CAR estimates from California high schools. The purpose of this paper is to examine and compare currently used methods of estimating CAR to present improved methods for CAR estimation.

Data and Methods Used to Evaluate Different Types of CAR Estimates

This investigation was conducted using information from the High School and Beyond Survey (HSAB) (Tourangeau, 1983). The HSAB is a major nationwide longitudinal study that is following 10,628 high school graduates of the class of 1980 and 10,934 graduates of the class of 1982. These graduates come from 923 high schools. Demographic information about the students (e.g., ethnic group, gender, etc.) and their high schools

---

Donald R. Hoover is an Assistant Professor in the Department of Epidemiology of The Johns Hopkins University School of Public Health, Baltimore, MD.

*Funding for this research was provided by the U. S. Army Grant Number DAFK 15-87-R-0166. The author thanks Dorothy Fidler for assistance and advice.

CG 022702

(e.g., size, location, etc.) was gathered. The college intentions of responding students were obtained during February of their senior year. The HSAB also recorded high school provided CAR estimates of their 1979 and 1981 classes and student college attendance after high school graduation.

The HSAB sample of American high school students was chosen using a two-stage process:

Phase I — High schools were selected nationwide. Over 72% of the schools initially selected participated. Those that chose not to participate were replaced by schools with similar characteristics.

Phase II — Students were selected from the schools chosen in Phase I. Over 80% of the students selected participated.

Due to the factors inherent in all large studies of this type, the HSAB sample was not completely random (Tourangeau, 1983). Certain school types (e.g., Catholic) and student types of special interest (e.g., low income white and high achieving black) were deliberately overselected. Students from certain demographic groups (e.g., low income and children of renters) were more likely either to choose not to participate or to later drop out of the study. Students selected from a given school in Phase II are not statistically independent of other students selected from the same school.

Non-random samples, such as this, are best analyzed using weighted replication methods given in Tourangeau (1983; SAS Institute, 1985). These procedures incorporate HSAB provided variables that identify the sampling subgroup of each student (and school) and the probability for that student (and school) of being in the final sample.

Analysis of Different CAR Prediction Methods

Three methods of predicting CAR will be explored and compared. The first method uses student intentions; the second uses educated guesses provided by high school officials; and the third method uses linear regression based on high school and student body characteristics. When evaluating CAR estimates, one should consider not only their precision and accuracy, but also the amount of effort needed to obtain those estimates.

Student Intentions

Questioning all the seniors at a given high school about their college intentions is an expensive and time consuming way to estimate a CAR. The HSAB questioned 1980 and 1982 seniors about their college intentions during February of their senior year. Responses were obtained from 89.7% of the 1980 graduates and 87.1% of the 1982 graduates and were categorized as follows:

A. Do not plan to attend college.
B. Unsure about future college plans.
C. Plan to attend college:
    1. After a year or more layoff following HS graduation.
    2. Part-time during the year following HS graduation.
    3. Full-time during the year following HS graduation.

Most schools requesting student intentions do not break category C into subcategories and hence it is not broken down in much of the analysis done here. The fraction and standard errors of the responding 1980 and

1982 graduates joining the CAR (attending college full-time during the year following high school graduation) by response category is given in Table 1.

Table 1
CAR by College Intentions for 1980 and 1982 Graduates

College Intention Category*	1980 Seniors Fraction in CAR	1980 Seniors Std Error of Fraction	1980 Seniors n	1982 Seniors Fraction in CAR	1982 Seniors Std Error of Fraction	1982 Seniors n	Average of 1980 and 1982 Fractions
A	0.050	0.0052	1494	0.031	0.0044	1556	0.040
A	0.115	0.0109	978	0.065	0.0073	1131	0.090
C (overall)	0.634	0.0052	7058	0.615	0.0043	7477	0.624
C.1	0.161	0.0121	877	0.107	0.0044	1088	0.134
C.2	0.389	0.0153	972	0.313	0.0155	894	0.351
C.3	0.769	0.0059	4951	0.762	0.0057	5495	0.765

*See text above for description

Student CAR by intention category differed between 1980 and 1982. The 1982 CAR fractions in each category were smaller than were the 1980 CAR fractions for the same category. Except for category C.3, the differences were statistically significant at $p<.05$. These differences could be due to the presence of a better economy and reduced college financial aid in 1982 than in 1980.

Table 1 shows that using student intentions directly to estimate CARs can produce inaccurate results since students will not always carry out their stated plans. For instance, if a school with 100 graduates finds that 70 plan to attend college, 20 do not plan to attend, and 10 do not know, it might estimate the CAR to be between 70% and 80%. In reality, if (A), (B) and (C) are the fractions of a high school's graduates in the corresponding college intention category of Table 1, then the expected CAR of a school is given by the following equation:

$$(\text{True CAR})_{H\,S} = 0.040\,(A) + 0.090\,(B) + 0.624\,(C).$$

This equation uses the averaged 1980 and 1982 college attendance fractions in Table 1 (column 8). Since (A) = .20, (B) = .10 and (C) = .70 for the above school, the estimated CAR percent for that school is 45.38%. The standard error of this CAR estimate is 4.24%. The above CAR estimate is somewhat imprecise as is indicated by the 95% confidence interval for the true CAR (36.9% < True CAR < 53.9%). It is still a great improvement over the unadjusted estimate of 70%-80%.

Graduates who indicated no plans to attend college, who planned to attend college after a year or more layoff following graduation, or who were unsure of their college plans, were very unlikely to become part of the CAR. On the other hand, those who indicated plans to attend college full-time the following year were likely, but by no means certain, to do so, which may suggest that counselors will have a more immediate effect attempting to help seniors with college plans to carry them out as opposed to encouraging students with no college plans to attend college.

High School Officials' Estimates

The easiest way to obtain a CAR estimate for a school is by asking the head guidance counselor or principal at the school. Obviously such estimates are subjective, and personal knowledge of the particular school official giving an estimate (which, of course, is not available in the HSAB data) could be useful in judging the bias of that estimate. The overall accuracy of such school-provided estimates were studied using HSAB data. The HSAB requested each high school to provide CAR estimates for their 1979 and 1981 graduating classes. Unfortunately, estimates of the 1980 and 1982 class CARs were not requested.

While a high school's CAR can fluctuate from year to year, much of this fluctuation will be due to randomness and other unmeasurable factors (Hoover, 1988). Hence, it is not unreasonable to assume that a school-provided estimate of its CAR for a given year (i.e., 1979) will be close to the CAR estimate it provides for the following year (i.e., 1980). Therefore, the 1979 (or 1981) school-provided CAR estimate can be compared with the 1980 (or 1982 respectively) CAR observed in the HSAB students sampled from that school.

The school-provided estimates differed from the actual school CARs. This is best shown by the following weighted simple linear regression giving the best unbiased estimators of 1980 (or 1982) high school CAR in terms of the school-supplied 1979 (or 1981) estimate.

$$(\text{True 198 CAR})_{\text{H.S.}} = 0.191 + 0.542 \cdot (\text{1979 School-Provided Estimate})_{\text{H.S.}}$$

$$(\text{True 1982 CAR})_{\text{H.S.}} = 0.159 + 0.548 \cdot (\text{1981 School-Provided Estimate})_{\text{H.S.}}$$

The differences between the parameter estimates of the two regressions are not statistically significant for $p<.05$. A good equation to use for adjusting a CAR estimate of a high school class provided by a school official is obtained by averaging the parameters of the previous two equations and is given below.

$$\text{True CAR})_{\text{H.S.}} = 0.175 + 0.545 \text{ (School-Provided CAR Estimate)}_{\text{H.S.}}$$

Using this equation, if a school-provided CAR estimate is 0%, then a better estimate woud be 100 [(0.175) + 0(.545)] = 17.5%. If a school-provided CAR estimate is 100%, then a better estimate would be 100 [(0.175) + 1.0 (0.545)] = 72.0%.

School officials from schools with low CARs tend to underestimate CAR and those from schools with high CARs tend to overestimate CAR. Part of this observed pattern could be an artifact caused by the use of CAR estimates for the previous year in the model and because it is the nature of linear regression to centralize extreme values. A larger part of this pattern probably results from school officials tending to give an all or none response when estimating CAR of students. That is, a counselor considering 100 students, each having a 70% chance of joining the CAR, may look at each student individually, decide that each student will join the CAR, and predict a group CAR of 100%, which is too high. Similarly, the counselor could underestimate the overall CAR for a group of students, each of whom has a low probability of joining the CAR.

A measure of the precision of the school-provided CAR estimate is the correlation coefficient between that estimate and the true high school CAR. The HSAB does not have true high school CARs because not all of the students from any high school were followed. A sample estimate

of the high school CAR can be calculated using the students from that school who were followed by the HSAB. The variability of this estimate due to sampling error can be calculated and adjusted for to obtain an approximate value for the above correlation coefficient as in Hoover (1988).

The correlation between the 1979 school CAR estimate and the 1980 high school CAR was approximately .473 while that between the 1981 school CAR estimate and the 1982 CAR was about .484. The fact that these are so low (a correlation of at least .70 is usually considered desirable) indicates that the school-provided CAR estimate are not precise. Nevertheless, school-provided CAR estimates adjusted by the method in this section are reasonably effective.

Linear Regression

Linear regression predicting high school CAR based on high school and student body characteristics (i.e., school ownership, ethnic composition. etc.) at first glance may appear to be a good objective way to estimate CAR. A linear predictive model for CAR would estimate CAR by: $CAR = B_0 + B_1 X_1 + \ldots + B_n X_n$ (where $X_1, X_2, \ldots, X_n$ are variables taken from the high school and $B_0, B_1 \ldots, B_n$ are parameters which are estimated from the HSAB data using weighted least squares). Such models, however, are cumbersome and difficult to work with.

Linear predictive models for the 1980 and 1982 high school CAR were developed. The predictor variables were chosen by variable selection from a larger set of variables in the HSAB data. Variables which would be too difficult to obtain in practice, such as student family income, were excluded from consideration. Table 2 gives the variables used, the 1980

Table 2
Linear Predictive Models

Variable	1980 Parameter Estimate	1982 Parameter Estimate	P-Values for Equality of 1980 & 1982 Parameters
Intercept	0.345	0.316	
CHS - Catholic High School	0.218	0.267	
PHS - Private Non-Catholic H. S.	0.213	0.218	
PAS - Proportion Asian Students	0.204	0.227	
SCZ - No. of Students in Senior Class	0.00015	0.000009	<0.05
PF - Proportion Female Students	0.038	0.044	
SNW - Public School with at least 60% Students Non-White	-0.056	-0.109	<0.05
HS - High School is in suburban area	0.030	0.056	
LD - Log_{10} of distance to nearest coll.	-019	-0.019	

CHS = (1 if school is Catholic, 0 otherwise)
PHS = (1 if school is private non-Catholic, 0 otherwise)
PAS = (Proportion of Asian students in the senior class)
SCZ = (Number of students in senior class)
PF - (Proportion of female students in the senior class)
SNW - (1 if school is public with at least 60% of senior class non-white, 0 otherwise)
HS - (1 if high school is in a suburban area, 0 otherwise)
LD - (Log_{10} [1 + distance from high school to nearest 2 or 4 year college])

and 1982 model parameter estimates, and the p-values for equality of the 1980 and 1982 parameters.

The linear models show interesting relationships between the CAR and the various factors. For instance, in 1982, given that all other factors were equal, a female CAR was 4.4% more than was a male's and a Catholic school graduate's CAR was 26.7% higher than was a public school graduate's.

The linear models did not, however, perform well in predicting CAR. First, the model did change over time. The coefficients of the variable SCZ (senior class size) and SNW (public non-white school) differed between the 1980 and 1982 models at $p<.05$. More restricted college financial aids and/or a better economy in 1982 than in 1980 could have caused relatively fewer students from non-white schools and larger inner city schools to attend college that year.

Second, and more important, the linear models were not precise. They explained only 28.7% and 29.2% of the variance in the 1980 and 1987 high school CAR respectively, compared to 47.3% and 48.4% respectively explained by models based on school-provided estimates. Without the Catholic and non-Catholic private school variables, the models will explain less than 9.5% and 9.6% of the 1980 and 1982 high school CAR variance respectively. Thus, they do a very poor job of distinguishing between CARs of public schools.

However, two variables not available in the HSAB data might have greatly improved the predictive ability of linear regression. They are average SAT (or ACT) score and percentage of graduating seniors taking the SAT (or ACT). This information can only be obtained from each individual high school and some high schools may not be willing to release it.

Summary

This study shows two methods can produce reasonably accurate CAR estimates; the first using educated guesses provided by high schools and the second using student intentions. A third method studied, linear regression with high school demographic characteristics, found interesting relationships to exist between the variables and CAR, but demonstrated poor predictive ability. The two ways of obtaining reasonable CAR estimates are summarized below.

An individual with little time or other resources could ask an appropriate school official to provide an educated guess of the CAR. This estimate should then be adjusted, using the third equation in the "High School Officials' Estimates" section, to eliminate bias caused by a tendency of officials to give extreme estimates. An individual estimating the CAR of a school that records the college intentions of its graduates could use these intentions to estimate CAR. The student college intention results should be adjusted by the methods in the "Student Intentions" section to compensate for individuals changing plans.

Discussion

It is difficult to directly compare student intention estimates to school official estimates since the overall distribution of student intentions for each high school is unknown. It is, however, probably best to use student

intention based estimates whenever possible, since they will not be influenced by the subjectivity of one person as are estimates provided by school officials.

The predictive ability of linear regression estimates of CAR could be greatly improved by using SAT (or ACT) test scores and percentages of the graduating class taking college entrance examinations as predictor variables. This information was not available in the HSAB study. A different type of predictive model, known as logistic regression, might also produce better results (SAS Institute, 1985). Currently available logistic regression techniques cannot handle the non-randomness of the HSAB sample.

It should be noted that significant numbers of students quit college before finishing, many during the freshman year (Tinto, 1987). Thus a large number of educators consider Freshman Year Completion Rate (FYCR), the proportion of a graduating high school class immediately enrolling full-time and completing one year of college, to be a better measure of college attendance than the CAR. Of the HSAB students joining the CAR, 87.5% completed the freshman year (Hoover, 1988). This was larger than a freshman year completion rate of 82.2% seen in a national cohort of 1972 college entrants by Eckland and Henderson (1981) and reported in Tinto (1987). The major reason for this difference is probably that the cohort followed here was required to be enrolled full-time to be part of the CAR, while Eckland and Henderson's cohort included part-time students who were more likely to drop out.

Completion of the freshman year among the CAR group was not numerically and statistically influenced by any of the factors considered in previous CAR models except that: (1) male students had a completion rate of 89.0% compared to only 86.2% for females; (2) Asian students had a very high completion rate of 96.2%; and (3) private (non-Catholic) high school students had a high completion rate of 95.2%.

One can use a CAR estimate of a graduating high school class given by any of the three methods discussed here to estimate the FYCR of that class. For public and Catholic high school classes consisting mostly of non-Asian student, the FYCR estimate is 0.875 time the CAR estimate. For private (non-Catholic) high schools, multiply the estimated CAR by 0.952 to obtain the FYCR estimate. For schools with predominantly Asian students, multiply the CAR estimate by 0.962 to obtain the FYCR estimate.

References

Eckland, B., & Henderson, L. (1981). *College attainment four years after high school.* Washington, D.C.: Center for Educational Statistics.

Hoover, D. (1988). *To college rate (TCR) and college drop rate (CDR) by high school predictive models series.* University of South Carolina Statistical Laboratory Technical Report 88-002.

Jones, C. (1986-a). *High school and beyond, 1980: Sophomore cohort second follow-up (1984.).* Washington, D.C.: Center for Educational Statistics.

Jones, C. (1886-b). *High school and beyond, 1980: Senior cohort second follow-up (1984).* Washington, D.C.: Center for Educational Statistics.

SAS Institute, Inc. (1985). *SAS users guides, version five.* Cary, NC.: Author.

Tinto, V. (1987). *Leaving college - Rethinking the causes and cures of student attrition.* Chicago, IL.: University of Chicago Press.

Tourangeau, R. (1983). *High school and beyond: First follow-up sample design.* Washington, DC.: Center for Educational Statistics.